

A PSYCHOMETRIC INVESTIGATION INTO THE USE OF AN ADAPTATION OF THE GHISELLI PREDICTABILITY INDEX IN PERSONNEL SELECTION¹

LIESLE TWIGGE

CALLIE THERON

*Department of Industrial Psychology
University of Stellenbosch*

HENRY STEEL

*Department of Psychology
University of Stellenbosch*

DEON MEIRING

SAPS

ABSTRACT

The magnitudes of validity coefficients typically encountered in validation studies are disappointingly low. Validity coefficients typically fall below 0,50 and only very seldom reach values as high as 0,70. Numerous possibilities have been considered on how to affect an increase in the magnitude of the validity coefficient. A thought-provoking alternative to the usual multiple-regression based attempts may be found in the work of Ghiselli (1956, 1960a, 1960b). The objective of this article is to propose and evaluate a modification to the original Ghiselli procedure. Encouragingly positive results were obtained. Recommendations for future research are made.

OPSOMMING

Die grootte-orde van geldigheidskoëffisiënte wat tipies in validasiestudies gevind word is teleurstellend laag. Geldigheidskoëffisiënte neem as 'n reël waardes kleiner as 0,50 aan en bereik by wyse van hoë uitsondering waardes so hoog soos 0,70. Verskeie moontlikhede in terme waarvan 'n verhoging in die geldigheidskoëffisiënt te weeg gebring sou kon word, is reeds oorweeg. 'n Stimulerende alternatief tot die gebruikelike meervoudige regressie gebaseerde pogings is te vind in die werk van Ghiselli (1956, 1960a, 1960b). Die doelstelling van hierdie artikel is om 'n wysiging aan die oorspronklike Ghiselli-prosedure voor te stel en te evalueer. Bemoedigend positiewe resultate is gevind. Aanbevelings vir verdere navorsing word gemaak.

Personnel selection strives to contribute to the organisation's overall efficiency by maximising the economic value added to the organisation by selecting those employees best suited for vacant positions (Boudreau, 1991; Cook, 1998; Guion, 1991). Personnel selection essentially is a form of applied decision-making. The focus thus should be on the quality of the selection decisions and not only on the psychometric properties of the measuring instruments used to provide the information for the decision-making. Cronbach and Gleser (1965) acknowledge the usefulness of tests for accurate estimation of an underlying latent variable, but suggest that the value of a selection procedure depends on many other qualities in addition to the reliability and validity with which the critical attributes are being measured. This should, however, not be interpreted to mean that (classical) measurement and test theory should be regarded as obsolete and irrelevant. Although it would be wrong to equate quality of decision-making to the magnitude of the validity coefficient, the latter nonetheless still influences the former. If the other pertinent factors affecting selection decision quality are held constant, selection decision quality increases as the absolute value of the validity coefficient increases. Utility is a positive linear function of validity, and for zero cost, is proportional to validity (Brogden, 1946; Brogden, 1949).

The validity coefficients typically encountered in validation studies are, however, disappointingly low. Validity coefficients typically fall below 0,50 and only very seldom reach values as high as 0,70 (Campbell, 1991; Guion, 1998). Typically selection instruments thus explain only 25% of the variance in the criterion (Campbell, 1991). The validity ceiling first identified by Hull (1928) seemingly still persists. Numerous possibilities have been considered on how to affect an increase in the magnitude of the validity coefficient (Campbell, 1991; Ghiselli, Campbell &

Zedeck, 1981; Guion, 1991; Guion, 1998; Wiggens, 1973). Most of these attempts revolved around modifications and/or extensions to the regression strategy (Gatewood & Feild, 1994).

A thought-provoking alternative to the usual multiple-regression based attempts may be found in the work of Ghiselli (1956, 1960a, 1960b). Rather than elaborating on the basic mathematical model of multiple-regression, Ghiselli has chosen to attack the problem of improved prediction directly by the use of empirical procedures (Ghiselli, 1956, 1960a, 1960b). The essence of the proposed procedure revolves around the development of a composite predictability index that explains variance in the prediction errors or residuals resulting from an existing prediction model. It would, however appear as if the procedure has found very little if any practical acceptance. The actuarial nature of the procedure could probably to a large extent account for it not being utilized in the practical development of selection procedures. The lack of general acceptance must, however, also be attributed in part to the fact that the predictability index originally proposed by Ghiselli (1956, 1960a, 1960b) failed to significantly explain unique variance in the criterion when added to a model already containing one or more predictors (Wiggens, 1973). The predictability index only serves the purpose of isolating a subset of individuals for whom the model provides relatively accurate criterion estimates. The selection problem, however, requires the assignment of each and every member of the total applicant sample (and not only a subset of the applicant group) to at least an accept or a reject treatment (Cronbach and Gleser, 1965) based on their estimated criterion performance.

Based on the original idea proposed by Ghiselli (1956, 1960a, 1960b), the objective of this research is to investigate the possibility that the differentiation between subjects on the basis of the predictability of their criterion performance

Requests for copies should be addressed to: CC Theron, Department of Industrial Psychology, University of Stellenbosch, Private Bag X1, Matieland, 7602

¹ The valuable comments and suggestions for improvement to this manuscript made by two anonymous reviewers are gratefully acknowledged.

could be used to increase the accuracy of the criterion estimates for the total applicant sample. More specifically, the objectives of the study are (a) to propose a modification to the Ghiselli procedure that would solve the aforementioned problem experienced by Ghiselli (1956, 1960a, 1960b) in his original studies, (b) to corroborate the earlier finding of Ghiselli (1956, 1960a, 1960b) that the development of a predictability index that significantly explains variance in the criterion residual is practically possible, (c) to demonstrate that the proposed modification to the Ghiselli procedure did in fact solve the problem experienced by the predictability index (based on absolute residuals) originally proposed by Ghiselli (1956, 1960a, 1960b), (d) to examine the factor structure of the modified predictability index to establish whether substantive theoretical meaning could be attached to it, (e) to examine the incremental validity resulting from the inclusion of the modified predictability index in the prediction model, and, (f) to examine the impact of the inclusion of the modified predictability index in the prediction model on selection utility.

Theoretical rationale for the development of a predictability index

Measurement data, once obtained, are translated into decisions in accordance with some strategy for decision-making (Cronbach, 1960). A decision strategy describes how scores from tests are to be combined with non-test information, and what decision will be made for any given combination of facts. A strategy is thus a rule for arriving at selection decisions used by a decision maker in any possible contingency (Cronbach & Gleser, 1965). It consists of a set of specified conditional probabilities (typically either zero or unity), which reflects the policy of decision-maker. In the final analysis it is the selection decision strategy that should be evaluated in terms of its predictive validity – in other words in terms of the correspondence that exists between the criterion referenced inferences made via the decision rule from the available predictor information and the actual criterion performance achieved (Gatewood & Feild, 1994).

Several selection decision-making strategies exist that range from pure clinical to pure mechanical combinations of data available to the decision maker (Gatewood & Feild, 1994; Grove & Meehl, 1996; Kleinmutz, 1990; Murphy & Davidshofer, 1988). Clinical prediction involves combining information from test scores and measures obtained from interviews and observations covertly in terms of an implicit combination rule imbedded in the mind of a clinician to arrive at a judgment about the expected criterion performance of the individual being assessed (Gatewood & Feild, 1994; Grove & Meehl, 1996; Murphy & Davidshofer, 1988). Mechanical prediction involves using the information overtly in terms of an explicit combination rule to arrive at a judgment about the expected criterion performance of the individual being assessed (Gatewood & Feild, 1994; Murphy & Davidshofer, 1988). An actuarial system of prediction represents a mechanical method of combining information to arrive at an overall inference about the expected criterion performance of an individual that was objectively derived via statistical or mathematical analysis from actual criterion and predictor data sets (Meehl, 1957; Murphy & Davidshofer, 1988). An actuarially derived decision rule should, therefore, more likely reflect the nature of the relationship that exists between the various predictor variables and the criterion construct. Regression analysis provides the basis of an actuarial decision-making strategy by regressing performance assessments on a weighted linear combination of predictors. The multiple regression strategy minimizes error in prediction and combines the predictors optimally to yield the most efficient estimate of criterion status (Berenson, Levine & Goldstein, 1983; Hair, Anderson, Tatham & Black, 1995; Gatewood & Feild, 1994; Howell, 1992).

The accuracy with which prediction models estimate criterion performance can be enhanced in a number of ways. Essentially two classes of approaches can be distinguished. The first category of approaches could be termed *substantive theory approaches* in as far as they originate from contemplating the manner in which variance in performance could be substantively explained in terms of theory. The second category of approaches could be termed *operational design approaches* in as far as they originate from reflecting on the degree of success with which the validation design measures the relevant latent variables and samples the relevant applicant population. The various arguments falling under these two categories of approaches essentially describe different but probably simultaneously operating processes that explain why existing prediction models make prediction errors and thus why the criterion performance of some individuals are predicted more accurately than the performance of others.

Under a substantive theory approach it would be argued that effective selection is possible because the performance level achieved by any individual on the job or in training is not a random event. There exists a systematic, albeit complex, relationship between specific person-centred characteristics, specific variables characterizing the job or training situation, and the level of success achieved on the job or in training. Effective selection is possible under a construct-orientated approach (Binning & Barrett, 1989) to the extent to which the identity of the person centred determinants of job or training performance are known and the manner in which they collectively combine in the criterion is accurately captured in a nomological network or latent structure (Campbell, 1991; Kerlinger & Lee, 2000). These person-centred determinants of criterion performance, could serve in combined form as a suitable substitute measure for the, still to be realised, actual criterion scores. The way measures of these determinants of performance should be combined is suggested by the way these determinants are linked in the nomological network (Theron, 1999). Typically the assumption is made that the linkages in the nomological network are linear. This need, however, not necessarily be the case.

To the extent that the linearity assumption is in error, the accuracy of prediction will suffer. To the extent that influential determinants of criterion performance are excluded from the prediction model, the accuracy of prediction will suffer. The accuracy with which prediction models estimate criterion performance can therefore be enhanced by building additional determinants of criterion performance into the model and/or by making provision for non-linearity in the model by including product or quadratic terms in the regression equation, which allows the model to remain linear in the partial regression coefficients (making provision for moderator variables would be a specific example of this strategy) or by formulating an equation which is non-linear in the regression coefficients (Berenson, Levine & Goldstein, 1983; Hair, Anderson, Tatham & Black, 1995; Gatewood & Feild, 1994; Howell, 1992).

The assumption that the relationship between the latent predictor variables and the latent criterion is linear (so as to simplify analysis) or at worst curvilinear, but expressible in terms of a familiar and solvable mathematical function could, however, still be insufficient to accurately model the relationship. If a highly contorted hyperplane defining the value of an endogenous criterion latent criterion variable (η) over a space of n exogenous latent predictor variables (ξ) would be assumed, such that for any combinations of conditions of the exogenous predictor variables the endogenous criterion latent variable has a specific value, the reaction of h to changes in ξ_i would seem random, even though h is strictly determined by ξ_i . One would thus have strict determinism masquerading as chaos so to speak (Theron, 2001). Should such a situation exist it would suggest the building of neural networks as the methodological avenue to pursue, rather than the conventional approach of fitting known, normally linear, mathematical

models, via regression analysis, to the data (Abdi, Valentin & Edelman, 1999; Anderson, 1995; Smith 1993).

An operational design approach, however, would attack the problem on how to enhance the accuracy with which prediction models estimate criterion performance differently. Under this approach the argument would be that when developing a selection procedure the objective is to model the relationship between the latent criterion construct and fallible measures of the predictor constructs that determine job performance as it exists in the applicant population on which the selection procedure will eventually be used. In reality, however, the relationship between a fallible measure of the criterion construct and fallible measures of the predictor constructs is modelled on a biased sample selected from the applicant population. The extent to which the operationalized criterion and/or the operationalized predictor contain systematic measurement error (i.e., bias) will distort the validity coefficient (Nunnally & Bernstein, 1994; Thorndike, 1982). The nature of the effect will depend on the patterns of correlations found between the contaminating variable, the predictor and the intermediate criterion. Hierarchical regression analysis, suppressor variables and partial correlation coefficients constitute options to address measurement bias, provided the source of the bias can be measured (Berenson, Levine & Goldstein, 1983; Hair, Anderson, Tatham & Black, 1995; Howell, 1992). The extent to which the operationalized criterion contains random measurement error and the extent to which the validation sample is a too homogenous and thus an unrepresentative, biased, sample from the applicant population, will adversely affect the validity coefficient (Campbell, 1991; Crocker & Algina, 1986; Lord & Novick, 1968; Messick, 1989; Schepers, 1996). Both of the latter factors will attenuate the validity coefficient. It thus follows that, to the extent that the aforementioned two factors did operate in the validation study but do not apply to the actual area of application, the obtained validity coefficient cannot, without formal consideration of these factors, be generalised to the actual area of application. The obtained validity coefficient thus cannot, without appropriate corrections, be considered an unbiased estimate of the actual validity coefficient of interest.

Appropriate formulas to correct the validity coefficient for criterion unreliability and restriction of range have been derived from classical measurement theory (Crocker & Algina, 1986; Lord & Novick, 1968; Kaplan & Saccuzzo, 2001; Schepers, 1996; Theron, 1999). If these corrections would be applied, the validity coefficient would be adjusted, but that would still leave the prediction equation, in terms of which the criterion estimates are derived, unaffected. The prediction equation actually used to derive the expected criterion estimates for decision-making is thus still the one derived from the validation study data, which, however, is not fully representative of the actual applicant population (Theron, 1999).

The approach suggested by Ghiselli (1956, 1960a, 1960b) seems to straddle the aforementioned two categories of approaches in terms of which the accuracy of prediction models can be enhanced. Classic psychometric theory holds that errors of measurement and of prediction are characteristics of the measuring device rather than the testee and that these errors are distributed randomly across individuals. Interactive effects between the measuring device and the person being assessed are not recognized, and the psychological structure of all individuals is taken to be the same. To increase reliability and validity of measurement, attention is then entirely focused on the improvement of measurement devices. However, a substantial body of evidence indicates there are systematic individual differences in error, and in the importance that a given trait has in determining a particular level of performance (Ghiselli, 1963).

Ghiselli (1960b) proposed a method whereby a moderator variable may be developed for a specific prediction situation.

Ghiselli (1956) investigated the possibility of differentiating by some other means, perhaps another test, those individuals whose predicted and actual criterion scores show small absolute discrepancies from those individuals whose predicted and actual criterion scores are markedly different. In a derivation sample, the absolute differences between predicted and actual criterion scores are obtained. Correlation analysis is subsequently performed to identify items from a separate item pool that discriminate between high and low predictability (i.e., items that correlate with the absolute differences between predicted and actual criterion scores). The items that correlate significantly with the absolute residual are then linearly combined in a predictability index. To the extent that the predictability index correlates with the absolute residuals, it should be possible to separate those subjects for whom the regression model provides accurate criterion estimates from those for whom the model performs less well. The index of predictability should therefore function as a moderator (Anastasi & Urbina, 1997; Wiggins, 1973). Knowledge of the predictability of an individual's criterion score should have considerable practical value. In an actual applicant sample, applicants would be ordered on the predictability index, and predictions would be made from the original predictors for the most predictable subset of applicants only. As predictions would be limited to an increasingly smaller proportion of the applicant sample, the validity of the predictor should approach unity. Selection procedures, therefore, can be improved not only by the addition of highly valid predictors to present procedures, but also by the addition of devices to screen out individuals whose levels of aptitude and job proficiency show little correspondence. Ghiselli (1956, 1960a, 1960b, 1963) has provided a number of convincing demonstrations of the utility of this approach and of variations on it (Wiggins, 1973).

However, it appears (Wiggins, 1973) that a combination of predictor and predictability index scores in multiple regression does not improve prediction over that given by the predictor scores alone. The value of predictability index scores lies solely in providing an index of the extent to which prediction of criterion scores from a particular test will be in error. The method does not provide for an alternative means of predicting those individuals who have been screened out because of their low predictability. Personnel selection, however requires that each and every applicant should be assigned to either an accept or a reject treatment (Cronbach & Gleser, 1965).

An important aspect in the original Ghiselli proposal that could hold the key to overcoming this shortcoming is the direction of the differences between actual and predicted scores of performance. Ghiselli viewed this as unimportant, as both over- and underestimates count as "errors" (Wiggins, 1973). However, the question arises whether the direction of the prediction error should not be taken into account when developing a predictability index? The addition of such an index to a selection battery could conceivably add to the predictive validity of the battery. What is required to improve predictive accuracy is the addition of a predictor to the regression model which functions by way of analogy like a an observation post adjusting the distance and angle of mortar or artillery fire onto a target. The predictors in the model provide criterion estimates that are in most cases too high or too low. If a predictive index could be developed which would provide feedback on the magnitude of the prediction error derived from the regression model as well as the direction of the error, then the inclusion of such an index in the regression equation as an additional main effect should logically enhance the predictive validity of the selection battery. This would, however, mean that the predictive index should be developed from the real differences between actual and predicted criterion scores of subjects, rather than the absolute difference as Ghiselli (1956, 1960a, 1960b, 1963) originally proposed. If the direction of the prediction error would be taken into account when developing a predictability index, large positive values on the index would signal large positive residuals (underestimation) and large negative values (or low positive

values) would signal large negative residuals (overestimation), assuming a positive correlation between the predictability index and the real residuals ($Y - E[Y|X_i]$).

The addition of this index to a regression model should enhance the predictive validity of the selection procedure because its values would provide feedback on the magnitude of the prediction error derived from the regression model as well as the direction of the error. The partial regression coefficient associated with the predictability index in the expanded regression model should be positive. An initial estimate derived from the original model, which is too low (underestimate) should therefore be elevated in the subsequent estimate derived from the expanded regression model due to the influence of the positive predictability index value. Conversely an initial estimate derived from the original model, which is too high should be lower in the subsequent estimate derived from the expanded regression model due to the influence of the negative predictability index value. The same principle should still apply even if the predictability index scale would be linearly transformed to run from zero to some positive upper limit.

The foregoing argument, however, still provides no substantive theoretical explanation as to why the proposed modification to the original Ghiselli procedure would assist in enhancing the predictive accuracy of an existing prediction model. The proposed modification to the original Ghiselli procedure is implicitly based on an argument as to why an existing prediction model predicts the criterion performance of some applicants more accurately than the performance of other applicants? Neither does the foregoing argument shed light on the related question why specific items would demonstrate the ability to reflect and even anticipate the prediction errors made by an existing prediction model? Systematic variance in the criterion is induced by systematic differences in a complex nomological network of person-centred and situational latent variables. Criterion performance is determined by the push and pull forces of a large number of variables. Criterion performance is a hyper plane responding to changes in p-1 performance determinants in a p dimensional space. To the extent that influential determinants of criterion performance are excluded from the prediction model, the accuracy of prediction will suffer because the push and/or pull effect of numerous influential variables on criterion performance is ignored. The extent to which prediction accuracy will suffer will, however, vary across individuals. For some individuals the omitted variables exerted a marked push or pull force to dramatically adjust the effect of the predictor(s) currently taken into account by the prediction model on criterion performance. For others the effect of the omitted variables on criterion performance is less dramatic. Could it be that the proposed modification to the original Ghiselli procedure essentially sniffs out item indicators of some of the latent variables that were not included in the prediction model but that do in fact influence performance?

Accuracy of prediction in and by itself is not the ultimate objective of research in personnel selection. The ultimate purpose of personnel testing is to arrive at substantiated qualitative decisions (Cronbach & Gleser, 1965). The challenge for any study into the improvement of personnel testing therefore ultimately lies in demonstrating that the quality of decision-making benefits from the proposed improvement. Several utility models can be distinguished to determine the total utility of a selection procedure, whereby the best known models are those of Taylor-Russell (1939), Naylor-Shine (1965), Brogden (1946) and Cronbach and Gleser (1965). Brogden (1946; 1949a; 1949b) and Cochran (1951) have shown that selection utility is a linear function of test validity, and that total selection utility could therefore be enhanced by an improvement in total validity. This increase in utility would in the final analysis determine whether the use of the proposed predictability index would contribute to the ultimate aim of effective selection in organisations, namely to contribute to the efficiency of the business in terms of monetary value.

METHOD

Participants

To serve the analytical purposes of this study, the data had to meet a number of specific requirements. The data set, firstly, had to contain an explicit criterion measure and a predictor measure, which correlates significantly with the criterion. The data set, secondly, had to contain the results of a second predictor, but in this case measures were required on the item level. The items of the second predictor had to provide the data from which the predictability indices would be harvested. No specific requirements were posed with regards to the nature of the latent variable measured by the second donor predictor. It was thus not required that the donor predictor measure should measure one or latent variable that could theoretically be expected to explain variance in the criterion construct. This rather liberal approach should, however, probably be questioned as somewhat naïve in as far as it completely ignored the question why specific items correlate with real or absolute residuals. The data set, thirdly, had to be large enough to allow the formation of a derivation sample on which the predictability index would initially be developed, and a holdout sample on which the predictability index would be cross-validated.

A data set was obtained from the data archives of Psytech SA that satisfied the first two of the aforementioned requirements. Psytech SA obtained data from the Gordon's Institute of Business (GIBS) on 101 MBA students between 1990 and 1991. A highly selected non-probability sample was chosen from students with average or above average interim MBA performance levels. The variance on the MBA examination scores was therefore typically low. Average interim MBA performance was utilized as the criterion in the study. The Ability, Processing of Information and Learning battery (Apil-B) (Taylor, 1994) was utilized as the predictor. Descriptive statistics on the criterion and the predictor is shown in Table 1. The Organisational Personality Profile (OPP) Questionnaire (Psytech, 2003), along with the Critical Reasoning Test Battery Version 2 (CRTB2) (Psytech, 2003) was also administered to the sample. The initial intention was to use only the items of the OPP for the development of the two predictability indices. It, however, subsequently become necessary to also use the items the CRTB2 for the development of the predictability index based on absolute residuals.

TABLE 1
DESCRIPTIVE STATISTICS ON THE APIL AND MBA PERFORMANCE DISTRIBUTIONS

		Apil general learning potential score	MBA Average to date
N	Valid	101	101
	Missing	0	0
Mean		63,162	67,861
Median		63,470	67,438
Mode		65,000	64,000
Std. Deviation		10,554	4,503
Variance		111,391	20,273
Skewness		-,359	,423
Std. Error of Skewness		,240	,240
Kurtosis		-,570	,055
Std. Error of Kurtosis		,476	,476
Minimum		37,000	58,750
Maximum		83,000	81,000

More detailed information regarding the sampling methodology was not available from Psytech. The nature of the sampling methodology is, however, not critical in arriving at valid and credible conclusions on the merits of the modifications proposed to the original Ghiselli procedure.

The data set obtained from Psytech was too small to permit the formation of a derivation sample and a holdout sample. In terms of Cohen's statistical power tables (Cohen, 1988), however, the sample size of 101 for the derivation sample can be regarded as adequate. The required number of participants to achieve statistical power of 0,80 in testing the significance of a sample product moment r , given a medium effect size of $r = 0.30$, a 5% significance level and a directional alternative hypothesis, is $n = 68$. At a 1% significance level the required n increases to 107. For a non-directional alternative hypothesis the Cohen tables recommend sample sizes of 84 ($p = 0,05$) and 124 ($p = 0,01$), assuming the same effect size as before.

Statistical hypotheses

Hypothesis 1: Average MBA performance (Y) is significantly related to learning potential as measured by the Apil-B (X_1).

$$H_{01}: \rho[Y, X_1] = 0$$

$$H_{a1}: \rho[Y, X_1] > 0$$

Hypothesis 2: A predictability index (X_2) can be developed from the items of a personality measure that shows a significant correlation with the real, algebraic residuals ($Y - E[Y|X_1]$) (Y_{res}) computed from the regression of the criterion (Y) on a learning potential predictor (X_1).

$$H_{02}: \rho[Y_{res}, X_2] = 0$$

$$H_{a2}: \rho[Y_{res}, X_2] > 0$$

Hypothesis 3: The addition of the predictability index, based on the real, algebraic values of the residuals (X_2), to the regression model will significantly explain unique variance in the criterion measure (Y) that is not explained by the learning potential predictor (X_1).

$$H_{03}: \beta_2[X_2] = 0 \mid \beta_1[X_1] \neq 0$$

$$H_{a3}: \beta_2[X_2] > 0 \mid \beta_1[X_1] \neq 0$$

Hypothesis 4: A predictability index (X_3) can be developed from the items of a personality measure that shows a significant correlation with the absolute residuals $|Y - E[Y|X_1]|$ ($|Y_{res}|$) computed from the regression of the criterion (Y) on a learning potential predictor (X_1).

$$H_{04}: \rho[|Y_{res}|, X_3] = 0$$

$$H_{a4}: \rho[|Y_{res}|, X_3] > 0$$

Hypothesis 5: The addition of the predictability index, based on the absolute values of the residuals (X_3), to the regression model will not significantly explain unique variance in the criterion measure (Y) that is not explained by the learning potential predictor (X_1).

$$H_{05}: b_2[X_3] = 0 \mid \beta_1[X_1] \neq 0$$

$$H_{a5}: b_2[X_3] > 0 \mid \beta_1[X_1] \neq 0$$

Postulate 1: The factor structure underlying the items comprising the predictability index (X_2) provides evidence that a clear substantive theoretical interpretation could be attached to the predictability index.

Postulate 2: If the addition of the predictability index, based on the real, algebraic values of the residuals (X_2), to the regression model significantly explains unique variance in the criterion measure (Y) that is not explained by the learning potential predictor (X_1) and thereby increases the predictive validity of the selection procedure, the addition of the predictability index, based on the real, algebraic values of the residuals (X_2) will increase selection utility.

Statistical analyses

The Statistical Package for Social Sciences (SPSS) version 11.0 was used to analyse the data. The specific analyses performed and the logic underlying the sequence of analyses will be outlined below.

RESULTS

To be able to investigate the feasibility of the proposed modifications to the original Ghiselli procedure, a significant linear relationship between a criterion and at least one predictor is required. It had been hypothesized that MBA performance should be systematically related to learning potential as measured by the Apil. Hypotheses 1 was tested by calculating the zero-order product-moment correlation between average MBA performance and performance on the Apil and the corresponding conditional probabilities $P[|r_{ij}| \geq r_c | H_0: \rho[Y, X_1] = 0]$. Given a 5% significance level and directional alternative hypotheses, H_{01} will be rejected if $P[|r_{ij}| \geq r_c | H_{01}: \rho[Y, X_1] = 0] < 0,05$. The matrix of zero-order Pearson correlation coefficients and the corresponding conditional probabilities is portrayed in Table 2.

The convention proposed by Guilford (cited in Tredoux & Durrheim, 2002, p. 184) has been used to interpret sample correlation coefficients. Although somewhat arbitrary and although it ignores the normative question about the magnitude of values typically encountered in a particular context, it nonetheless fosters consistency in interpretation.

The moderate positive correlation of the Apil-B ability test (X_1) and the MBA performance (Y) ($r = 0,46$; $p < 0,05$) confirmed that the Apil-B can be used as the primary predictor of MBA performance. H_{01} can therefore be rejected. The substantial relationship between learning potential and MBA performance can thus be used as a platform to empirically investigate the proposed modifications to the original Ghiselli procedure.

TABLE 2
CORRELATION BETWEEN THE APIL-B ABILITY TEST (X_1) AND
MBA PERFORMANCE (Y) (N= 101)

		MBA Average to date (Y)	Apil general learning potential score (X_1)
MBA Average to date (Y)	Pearson Correlation Sig. (1-tailed)	1 .	0,416 ,000
Apil general learning potential score (X_1)	Pearson Correlation Sig. (1-tailed)	0,416 ,000	1 .

Average MBA performance was subsequently regressed on the Apil-B ability test (X_1) by fitting the following regression model on the data:

$$E(Y|X_1) = \alpha + \beta[X_1].$$

The results of the standard regression analysis are presented in Table 3. Approximately 17% of the variance in the criterion (MBA performance) can be explained in terms of performance on the Apil-B (the primary predictor).

TABLE 3
SIMPLE LINEAR REGRESSION OF AVERAGE MBA
PERFORMANCE ON LEARNING POTENTIAL

R	R Square	Adjusted R Square	Std. Error of the Estimate		
0,416	0,173	0,164	4,115620		
	Sum of Squares	df	Mean Square	F	Sig.
Regression	350,366	1	350,366	20,685	0,000
Residual	1676,895	99	16,938		
Total	2027,261	100			
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
B		Std. Error	Beta		
(Constant)	56,660	2,497		22,693	0,000
Apil general learning potential score	0,177	0,039	0,416	4,548	0,000

The real, algebraic unstandardized residuals ($Y - E[Y|X_1]$) and the absolute unstandardized residuals ($|Y - E[Y|X_1]|$) were subsequently derived from the fitted regression model and written to the active data file. The real, algebraic unstandardized residuals are plotted against the predictor in Figure 1. From Figure 1 it appears as if the linearity, normality and homoscedasticity assumption underlying the linear model have been reasonably well satisfied (Tabachnick & Fidell, 1989). Satisfaction of the homoscedasticity assumption would, moreover, imply that accuracy of prediction is not a function of learning potential. Accuracy of prediction is, however, a (linear) function of criterion performance, with the strength of the relationship inversely related to the predictive validity of the predictor. Large positive real residuals tend to be associated with high MBA averages while high negative real residuals are associated with low MBA averages (not shown). Knowing this, however, has very little practical value in improving prediction accuracy other than to underline the need to increase predictive validity. The absolute unstandardized residuals are plotted against the predictor in Figure 2.

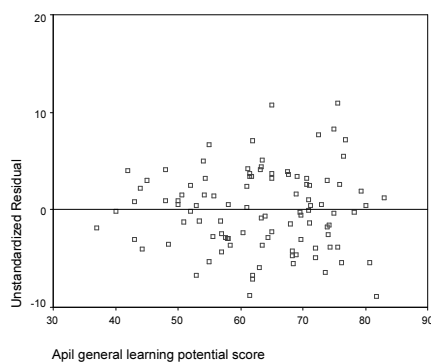


Figure 1: Real, algebraic unstandardized residuals plotted against learning potential

Descriptive statistics for the real and absolute unstandardized residuals are shown in Table 4. In the case of the real residuals, the skewness and kurtosis statistics do not deviate significantly ($p > 0,05$) from zero, thus supporting the inferences made from Figure 1.

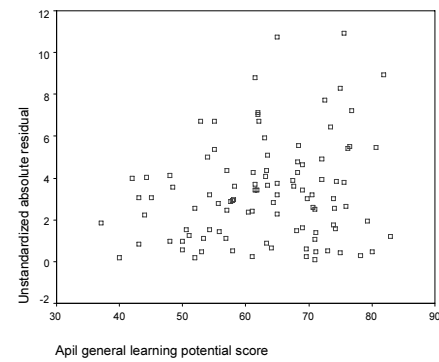


Figure 2: Absolute unstandardized residuals plotted against learning potential

TABLE 4
DESCRIPTIVE STATISTICS FOR REAL AND ABSOLUTE UNSTANDARDISED RESIDUALS

		Unstandardized real residuals	Unstandardized absolute residuals
N	Valid	101	101
	Missing	0	0
Mean		0,000	3,307
Median		-,064	3,016
Std. Deviation		4,095	2,393
Variance		16,769	5,724
Skewness		0,207	0,955
Std. Error of Skewness		0,240	0,240
Kurtosis		-0,118	0,853
Std. Error of Kurtosis		0,476	0,476
Minimum		-8,922	0,064
Maximum		10,943	10,943

The 98 individual items of the OPP personality questionnaire were subsequently correlated with the real and absolute residuals computed from the fitted regression model. The OPP items that correlated significantly with the real residuals at the 0,05 level were flagged for inclusion in the predictability index (X_2). Nine items correlated significantly with the real residuals at this level (minimum $r = 0,196$; maximum $r = 0,315$; average $r = 0,220$). In the case of the absolute residuals, however, only a single OPP item presented itself as a significant predictor of the absolute prediction errors made by the fitted regression model. This clearly created a dilemma as far as the calculation of the second predictability index (X_3) is concerned. The possibility of harvesting items from the Critical Reasoning Test Battery (CRTB2) was consequently examined. The 62 items of the CRTB2 subtests were therefore correlated with the absolute residuals in a similar fashion to the OPP items. Again the yield was rather disappointing. Only three CRTB2 items correlated significantly with the absolute residuals at the 0,05 level; two items from the Verbal subscale and one item from the Numerical subscale (minimum $r = 0,208$; maximum $r = 0,388$; average $r = 0,329$). It is worthy of note that the CRTB2 items yielded eight significant predictors of the real residuals (minimum $r = 0,245$; maximum $r = 0,362$; average $r = 0,273$). A further sobering fact is that although the number of items in the OPP and the CRTB2 that correlate significantly with the real residuals exceeded the number of significant correlations one could expect by chance on a 0,05 significance level (4,9 and 3,1 respectively), this is not the case with regards to the absolute residuals. Since

approximately five of the nine items harvested from the OPP could have been selected by chance alone, the danger exists that the predictability index took advantage of idiosyncrasies in the specific data set that would unlikely repeat itself a subsequent samples taken from the same population². The likelihood that the predictability index would cross-validate successfully thus diminishes.

The selected nine OPP items correlating with the real residuals were subsequently combined in an unweighted linear composite by taking the mean of the qualifying items, to form the predictability index (X_2) based on real residuals. The selected three CRTB2 items were likewise combined in an unweighted linear composite by taking the mean of the qualifying items, to form the predictability index (X_3) based on absolute residuals. The eight CRTB2 items significantly correlating with the real residuals and the single OPP item correlating significantly with the absolute residuals could also have been utilized in the formation of X_2 and X_3 respectively. It was, however, decided to restrict the harvesting of items to a single donor instrument so as to not run the risk of uncovering an obvious underlying factor structure reflecting nothing more than the nature of the instruments contributing items to the index when investigating postulate 1.

The predictability index based on the real residuals (X_2) and the predictability index based on the absolute residuals (X_3) were subsequently correlated with the unstandardized real and absolute residuals to determine the success with which the two predictability indices have been developed. In anticipation of the addition of the predictability indices to the basic regression model, the correlation of the two indices with the primary predictor and with the criterion was determined as well. The results are presented in Table 5.

Table 5 shows that the predictability index based on real residuals, (X_2), did correlate moderately (0,509) and significantly ($p < 0,05$) with the real residuals derived from the regressing the MBA averages on the Apil-B ability predictor. H_{02} can therefore be rejected in favour of H_{a2} . It is possible to develop a predictability index (X_2) from the items of a personality measure that shows a significant correlation with the real, algebraic residuals ($Y - E[Y|X_1]$) computed from the regression of the criterion on a learning potential predictor. Table 5, in addition, reveals that the absolute residual predictability index based on the absolute residuals (X_3) did correlate moderately (0,508) and significantly ($p < 0,05$) with the absolute residuals. H_{04} can therefore be rejected in favour of H_{a4} , if the initial assumption that the OPP would yield a

sufficient number of items for the index could be warranted. It is possible to develop a predictability index (X_3) from the items of a critical reasoning measure that shows a significant correlation with the absolute residuals ($|Y - E[Y|X_1]|$) computed from the regression of the criterion on a learning potential predictor.

As expected, the predictability index based on real residuals (X_2), correlated low (-0,002) and insignificantly ($p > 0,05$) with the absolute residuals derived from regressing the MBA averages on the Apil-B ability predictor. Likewise the predictability index based on absolute residuals (X_3), correlated low (-0,047) and insignificantly ($p > 0,05$) with the real residuals. Table 5, furthermore, suggests that that the inclusion of X_2 alongside X_1 in a multiple regression model is more likely to be meaningful than the addition of X_3 to a regression model already including X_1 . X_2 correlated low (0,056) and insignificantly ($p > 0,05$) with the Apil-B results while correlating moderately (0,487) with the criterion. The predictability index based on real residuals (X_2) therefore seems to explain unique variance in the criterion not explained by the primary predictor. X_3 correlates low (0,242) but statistically significantly ($p < 0,05$) with the predictor while correlating low (0,058) and statistically insignificantly ($p > 0,05$) with the criterion. The predictability index based on absolute residuals (X_3) therefore seems not to explain unique variance in the criterion.

Table 5 indicates that the unstandardized real residuals correlate very high (0,909) and statistically significantly ($p < 0,05$) with the MBA average. This could be interpreted to mean that the real residual and the criterion is essentially the same variable. Since the modified predictability index is constructed from items correlating with the real residual, one could argue that the whole exercise essentially boils down to using a variable to predict itself. This line of reasoning, however, ignores the fact that the total criterion sum of squares ($\Sigma(Y_i - E[Y])^2$) can be partitioned into a sum of squares due to regression ($\Sigma(E[Y|X_1] - E[Y])^2$) and a residual sum of squares ($\Sigma(Y_i - E[Y|X_1])^2$). The total variance can thus be partitioned into a proportion criterion variance that can be explained in terms of the Apil-B (0,416²) and a proportion criterion variance that cannot be explained in terms of the Apil_B (1-0,416²). The very high correlation observed between MBA average and the real residual is therefore simply an alternative expression of the fact that Apil_B only explains a small proportion (0,416² = 0,173) of the variance in MBA average performance. The remaining proportion of the variance in MBA average performance (0,909² = 0,827) is explained by the real residual.

TABLE 5
CORRELATIONS BETWEEN THE PREDICTABILITY INDICES, THE PRIMARY PREDICTOR AND THE CRITERION (N = 101)

		X_2	X_3	Unstandardized real residual	Unstandardized absolute residual	Apil general learning potential score (X_1)	MBA Average to date (Y)
X2	Pearson CorrelationSig. (2-tailed)	1.	-0,028 0,778	0,509 0,000	-0,002 0,984	0,056 0,576	0,487 0,000
X3	Pearson CorrelationSig. (2-tailed)	-0,028 0,778	1 .	-0,047 0,641	0,508 0,000	0,242 0,015	0,058 0,565
Unstandardized real residual	Pearson CorrelationSig. (2-tailed)	0,509 0,000	-0,047 0,641	1 .	0,075 0,456	0,000 1,000	0,909 0,000
Unstandardized absolute residual	Pearson CorrelationSig. (2-tailed)	-0,002 0,984	0,508 0,000	0,075 0,456	1 .	0,190 0,057	0,147 0,142
Apil general learning potential score (X_1)	Pearson CorrelationSig. (2-tailed)	0,056 0,576	0,242 0,015	0,000 1,000	0,190 0,057	1 .	0,416 0,000
MBA Average to date (Y)	Pearson CorrelationSig. (2-tailed)	0,487 0,000	0,058 0,565	0,909 0,000	0,147 0,142	0,416 0,000	1 .

² This important and highly relevant point was raised by an anonymous reviewer.

Table 5 finally also indicates that learning potential is not related to the accuracy of prediction (0,000; $p > 0,05$). This is also graphically portrayed in Figure 1 through the rectangular spread of real residuals across the range of Apil-B scores observed.

Descriptive statistics for the two predictability indices are provided in Table 6. Two dummy variables (X_2D and X_3D) were subsequently created by dichotomising the index distributions into high and low prediction accuracy groups. Since X_2 reflects the magnitude and direction of prediction error (i.e., real residuals), a low prediction error group, centred on zero had to be isolated. X_3 , in contrast reflect only the magnitude of prediction error and thus to isolate a high prediction accuracy, the cases falling below the median were flagged. On X_2 the cases falling between the twenty-fifth and seventy-fifth percentiles were classified as high prediction accuracy cases. On X_3 cases with an index score on or below the fiftieth percentile were classified as high prediction accuracy cases.

TABLE 6
DESCRIPTIVE STATISTICS FOR THE TWO PREDICTABILITY INDICES

		X_2	X_3
N	Valid	101	101
	Missing	0	0
Mean		3,0088	2,4422
Median		3,0000	2,5000
Std. Deviation		0,44905	0,46811
Variance		0,20165	0,21913
Skewness		-0,038	0,372
Std. Error of Skewness		0,240	0,240
Kurtosis		-0,363	0,171
Std. Error of Kurtosis		0,476	0,476
Percentiles	25	2,6667	2,0000
	50	3,0000	2,5000
	75	3,3333	3,0000

The relationship between the criterion and the predictor was subsequently graphically portrayed in Figure 3 and Figure 4 for the two levels of the dummy variable separately.

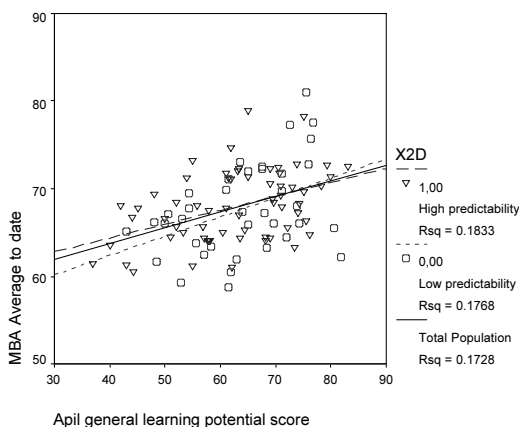


Figure 3: MBA average performance as a function of learning potential depicted for high ($X_2D=1$) and low predictability ($X_2D=0$) groups separately (predictability index based on real residuals).

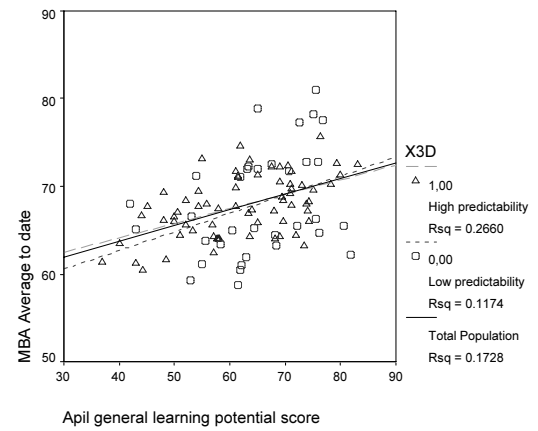


Figure 4: MBA average performance as a function of learning potential depicted for high ($X_3D=1$) and low predictability ($X_3D=0$) groups separately (predictability index based on real residuals).

Figures 3 and 4 seem to suggest that the predictability index based on the absolute residuals (X_3) is more effective in isolating a subset of individuals for whom the model provides more accurate criterion estimates than the predictability index based on real residuals (X_2). The two indices both correlate moderately strongly (0,51) with the residuals from which it is derived. The superiority of one index over the other in separating the more accurately predictables from the less accurately predictables thus is somewhat surprising.

Table 7 reveals that the addition of the predictability index, based on the real values of the residuals (X_2), to the basic regression model significantly ($p < 0,05$) explains unique variance in the criterion measure that is not explained by the learning potential predictor. H_{03} can thus be rejected in favour of H_{a3} . The original predictor still significantly ($p < 0,05$) explains variance in the criterion not explained by the predictability index. The expanded regression model explains approximately 39% of the variance in the criterion, compared to the approximately 17% explained by the basic model. The addition of the predictability index thus affected a substantial increase in the proportion of criterion variance explained.

TABLE 7
STANDARD MULTIPLE REGRESSION OF MBA PERFORMANCE ON LEARNING POTENTIAL AND THE PREDICTABILITY INDEX DERIVED FROM REAL RESIDUALS (X_2)

R	R Square	Adjusted R Square	Std. Error of the Estimate		
0,623	0,388	0,376	3,557492		
	Sum of Squares	df	Mean Square	F	Sig.
Regression	786,998	2	393,499	31,092	0,000
Residual	1240,263	98	12,656		
Total	2027,261	100			
	Unstandard- ized Coefficients	Standard- ized Coefficients	t	Sig.	Correlations
	B	Std. Error	Beta		Zero-order Partial Part
(Constant)	43,342	3,130		13,846	0,000
Apil general learning potential score (X ₁)	0,166	0,034	0,390	4,922	0,000 0,416 0,445 0,389
X ₂	4,661	0,793	0,465	5,874	0,000 ,487 0,510 0,464

Table 7 reveals that the unique variance in the predictability index (X_2) explains approximately 26% (0,510²) of the unique variance in the criterion after controlling for variance due to the Apil. The unique variance in the predictability index (X_2) explains approximately 22% (0,464²) of the total variance in the criterion. Judged by the standardized partial regression coefficients and the partial and semi-partial correlation coefficients the predictability index is the more influential predictor in the regression model. No convincing substantial theoretical explanation for this finding could be offered.

Table 8 reveals that the addition of the predictability index, based on the absolute values of the residuals (X_3), to the basic regression model does not significantly ($p > 0,05$) explain unique variance in the criterion measure that is not explained by the learning potential predictor. H_{05} can thus not be rejected in favour of H_{a5} . It could, however, be contended that the analysis is inappropriate in as far as an X_3 learning potential interaction effect should have been added to the model rather than an index main effect. Although no supporting evidence is presented here, this study also finds that the addition of a term representing the interaction between X_3 and Apil, also does not significantly ($p > 0,05$) explain unique variance in the criterion measure that is not explained by the learning potential predictor.

TABLE 8
STANDARD MULTIPLE REGRESSION OF MBA PERFORMANCE
ON LEARNING POTENTIAL AND THE PREDICTABILITY FROM
ABSOLUTE RESIDUALS (X_3)

R	R Square	Adjusted R Square	Std. Error of the Estimate					
0,418	0,175	0,158	4,131730					
	Sum of Squares	df	Mean Square	F	Sig.			
Regression	354,284	2	177,142	10,377	0,000			
Residual	1672,977	98	17,071					
Total	2027,261	100						
	Unstandard- ized Coefficients	Standard- ized Coefficients	t	Sig.	Correlations			
	B	Std. Error	Beta		Zero- order	Partial	Part	
(Constant)	57,429	2,977		19,293	,000			
Apil general learning potential score (X_1)	0,182	0,040	0,427	4,512	0,000	0,416	0,415	0,414
X_3	-0,436	0,910	-0,045	-0,479	,633	0,058	-0,048	-0,044

Given that the addition of the predictability index, based on the real values of the residuals (X_2) to the basic regression model significantly explains unique variance in the criterion measure that is not explained by the learning potential predictor (X_1), the question arises whether substantive meaning could be attached to the index scores. The objective was to determine if any theoretical meaning could be attached to the common factors underlying the index, if any were identified, and whether these interpretations would make sense in terms of the criterion. To shed light on this matter an exploratory principle component analysis was performed on the OPP items combined in the predictability index. The rotated component matrix should indicate whether the items comprising the predictability index systematically measured one or more underlying common construct(s), which could be linked to specific personality construct(s) or whether the predictability index is nothing more than an incoherent, meaningless collection of items that have nothing more in common than their correlation with the

regression residuals. The eigenvalue greater than one rule was used to decide on the number of factors to extract. Varimax rotation was used to rotate the obtained solution to simple structure.

Based on the eigenvalue greater than one rule and the scree plot four factors were extracted and orthogonally rotated (Table 9). The first four factors account for approximately 63% of the variance in the items. These results, however, fail to provide a clear, convincing, and credible answer to the question whether substantive meaning could be attached to the index scores. The borderline Kaiser-Maier-Olkin measure of sampling adequacy value (0,552) casts some doubt on the factorability of the correlation matrix (Tabachnick & Fidell, 1989). Extracting this many factors from only nine items and a sample size of 101 also seem somewhat questionable, especially given the unconvincing KMO statistic. No clear-cut picture moreover emerges from Table 9. Although each item loads reasonably high on single factor only, the common theme amongst the items loading on the same factor tends to be somewhat debatable. The first principle component could possibly be interpreted as a focus-intensity factor, the second principle component possibly as a compulsiveness factor and the third principle component possibly as a driven factor. These suggestions are, however, at best tenuous. Despite their questionable nature, these themes could conceivably play a role in the level of performance MBA students achieve. With the wisdom of hindsight this could, however, probably have been said for any of the OPP items. It should finally be conceded that it probably would have been more appropriate to have performed a common factor analysis rather than principal component analysis, given the intention to identify common factors. The nature of the pattern matrix obtained through principal axis factor analysis with oblique rotation roughly replicates the structure obtained through the principal component analysis, though somewhat less clean-cut. The small entries in the factor correlation matrix ($< |0,20|$) suggest that a single second-order factor is highly unlikely. The available evidence thus seems to suggest that the items combined in the predictability do not reflect a single underlying factor but fails to convincingly rule out the possibility that the predictability index is very little more than an incoherent, meaningless collection of items that have nothing more in common than their (possibly chance) correlation with the regression residuals. The most prudent option would probably be to regard the available evidence as too ambivalent to take any definite decision on postulate 1.

TABLE 9
ROTATED FACTOR MIX

Item	Component			
	1	2	3	4
I rarely have time for lunch.	-8,236E-02	7,373E-02	0,797	1,584E-02
I feel uncomfortable in crowded spaces (e.g. tube trains, lifts etc.).	-0,353	0,613	-7,243E-02	0,122
If I am near a friend's house I will often drop in just to say hello.	4,831E-02	0,677	0,157	0,300
Cleanliness is the greatest of all virtues.	0,107	0,799	7,285E-02	-0,334
I often have difficulty remembering things.	0,667	-7,153E-02	2,018E-02	4,068E-02
There never seems to be enough hours in the day to get everything done.	0,147	3,563E-02	0,809	3,571E-02
I am inclined to get tense before important meetings, particularly if much is at stake.	0,735	6,081E-02	0,323	-0,114
People are fundamentally goodhearted and kind.	-1,017E-02	6,081E-02	3,826E-02	0,937
I find it easy to persuade people of my point of view.	0,706	-3,181E-02	-0,145	1,260E-02

Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 5 iterations.

Item analyses were nonetheless performed on the set of nine items derived from the correlation between the OPP personality measurement and the real residuals taking into consideration the results of the principle component analysis. The results of the item analyses ($\alpha(C_1) = 0,5241$ $\alpha(C_2) = 0,4919$ $\alpha(C_3) = 0,5289$) indicate modest internal consistency for the three sets of items loading on the first three principle components. This finding is, however, not surprising given the limited number of items involved. Given the findings on the underlying structure it would not be meaningful to directly calculate a coefficient alpha for the nine items combined in the predictability index. The reliability of an unweighted linear composite (Nunnally & Bernstein, 1994) comprising the eight items loading on the first three principle components could be calculated though from the reliabilities and the variances of the three components. As could be expected a rather modest value of 0,601 is obtained.

A definite increase in the proportion of criterion variance explained was found when adding the predictability index based on real residuals to the basic regression model. The question is what the effect of this increase in predictive validity is on the quality of selection decision-making. The Taylor-Russell (Cascio, 1991), Naylor-Shine (Cascio, 1991) and Brogden-Cronbach-Gleser (Brogden, 1949; Cascio, 1991; Cronbach & Gleser, 1965) utility models were subsequently employed to describe the effect of the incremental validity of the predictive index on the quality of selection decision-making.

The addition of the predictability index resulted in an increase in predictive validity from 0,416 (Table 5) to 0,623 (Table 7). To translate this increase in predictive validity to increases in decision quality in terms of the aforementioned three utility models, however, requires additional data on the other selection parameters characterizing the three models. Since such data was not available for the validation sample, realistic illustrative values had to be assumed for the other parameters affecting the improvement in the quality of selection decision-making in each of the utility models to describe the effect of the incremental validity of the predictive index on the quality of selection decision-making. The choice of specific parameter values was essentially an arbitrary one. An applicant pool of 2000 and 100 vacancies was consequently assumed. Average tenure was assumed to be 5 years. The per-applicant cost associated with the Apil battery was assumed to be R250 and that of the OPP, R350. The standard deviation of the criterion distribution expressed in a R-c metric was assumed to vary between 35% and 45% of average salary (Cascio, 1991). Average salary was arbitrarily set at R100 000 per annum. It was assumed that 50% of the applicant pool could succeed if selected. Bivariate normality was assumed. The selection ratio ϕ would therefore equal 0,05 and the resulting λ value, obtained from the standardised normal probability table would equal 0,103. The base rate (BR) would be 0,50.

The improvement in the proportion of the selected applicants succeeding on the criterion (i.e., the success ratio, S_v) affected by the inclusion of the predictability index in the regression model, would under the aforementioned assumptions be given by equation 1:

$$\begin{aligned}\Delta S_v &= (S_v[X_1, X_2] - BR) - (S_v[X_1] - BR) \\ &= S_v[X_1, X_2] - S_v[X_1] \\ &= 0,9434 - 0,82388 \\ &= 0,11952 \text{-----} 1\end{aligned}$$

$S_v[X_1, X_2]$ and $S_v[X_1]$ were calculated via SPSS by calculating $P[Z_y^{30} \text{ and } Z_x^{31,64485}] / P[Z_x^{31,64485}]$ for the two validity coefficients, assuming multivariate normality. The addition of the predictability index (X_2) to the basic regression model would therefore, under the abovementioned scenario, result in an approximate 12% increase in the percentage selectees successful. This percentage would increase if larger increases in the validity coefficient could be affected.

The improvement in the mean standardized criterion performance of the selected group affected by the inclusion of the predictability index in the regression model, assuming a selection ratio of $\phi = 0,05$, will under the abovementioned scenario be given (in standard deviation units) by equation 2:

$$\begin{aligned}\Delta E[Z_y | \text{selected}] &= [R(Y, E[Y | X_1, X_2]) (\lambda / \phi)] - [r(Y, X_1) (\lambda / \phi)] \\ &= [R(Y, E[Y | X_1, X_2]) - r(Y, X_1)] (\lambda / \phi) \\ &= [0,623 - 0,416] [0,103 / 0,05] \\ &= 0,42642 \text{-----} 2\end{aligned}$$

The addition of the predictability index (X_2) to the basic regression model would therefore, under the abovementioned scenario, result in an increase in average performance of approximately 0,43 standard deviation units. This might seem rather trivial but when extrapolated over selectees, time periods, and when multiplied by the performance unit value of one standard deviation, could amount to an impressive quantity.

The R-c value of the improvement in the mean standardized criterion performance of the selected group affected by the addition of the predictability index to the basic regression model is be given (in R-c) by equation 3:

$$\begin{aligned}\Delta U &= TN_s R(Y, E[Y | X_1, X_2]) SD_y (\lambda / \phi) - (C_1 + C_2) N_a - \\ &\quad TN_s r(Y, X_1) SD_y (\lambda / \phi) - C_1 N_a \\ &= TN_s SD_y (\lambda / \phi) (R(Y, E[Y | X_1, X_2]) - r(Y, X_1)) - N_a (C_2 + 2C_1) \\ &= 5 [100] [40000] [0,103 / 0,05] (0,623 - 0,416) - 100 [500 + \\ &\quad 350] \\ &= R8 443 400-00 \text{-----} 3\end{aligned}$$

Where:

ΔU = the increase in utility due to the addition of the predictability index; T = the average predicted tenure of the selected applicants; N_s = the number of people selected for a position using a selection battery to which the index computed in the study has been added; $R(Y, E[Y | X_1, X_2])$ = the correlation coefficient obtained by adding the index to a selection battery already containing the ability predictor; SD_y = the standard deviation of the criterion distribution expressed in a R-c metric; λ = the height of the ordinate cutting off an area under the standardised normal distribution corresponding to a selection ratio ϕ ; ϕ = the selection ratio; C_1 = the per applicant cost for the Apil; $r(Y, X_1)$ = the validity coefficient of the basic regression model; and C_2 = the per applicant cost of the OPP.

The addition of the predictability index (X_2) to the basic regression model would therefore, under the abovementioned scenario, result in an increase in average performance worth R8 443 400-00 over the average tenure of 5 years. This is a somewhat overoptimistic estimate in as far as it fails to reflect the time value of future earnings and the tax liability higher performance earnings would imply (Cascio, 1991). The estimate, in conjunction with the other two utility estimates, nonetheless provides support for postulate 2.

To illustrate the linear relationship between the increase in validity affected by the predictability index and utility, equation 3 has been solved for a range of possible values for SD_y and $R(Y, E[Y|X_1, X_2])$, while fixing the remaining utility parameters at their initially chosen values. Schmidt and Hunter's (in Cascio, 1991) estimate of the standard deviation of the criterion distribution expressed in a R-c metric as 40 % of annual salary was varied with five percent up and down, resulting in the use of three values, i.e. 35%, 40% and 45%. The value of $R(Y, E[Y|X_1, X_2])$ was essentially varied in steps of 0,10 (see Table 10 and Figure 5).

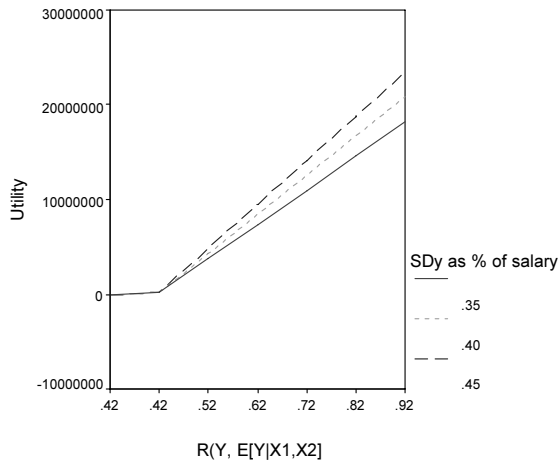


Figure 5: Incremental utility as a function of $R(Y, E[Y|X_1, X_2])$ and SD_y

Figure 5 illustrates the resultant increase in the monetary utility as the correlation coefficient $R(Y, E[Y|X_1, X_2])$ increased from 0,416, as well the acceleration in the increase in the utility when the standard deviation of the criterion distribution expressed in a R-c metric increased from 35% of annual salary to 40% to 45%.

DISCUSSION

The main findings of this study regarding the development of a predictability index are fourfold. It is possible to develop a predictability index, which correlates with the real, algebraic residuals derived from the regression of a criterion on one or more predictors. The addition of such a predictability index to the original regression model can produce a significant increase in the correlation between the selection battery and the criterion. This increase can trigger a substantial and useful increase in the utility of the selection battery. The potential benefits especially apply to companies selecting large numbers of employees per year at small selection ratios from even larger applicant pools. Although it is possible to develop a predictability index, which correlates with the absolute residuals derived from the regression of a criterion on one or more predictors, the addition of such a predictability index to the original regression model does not produce a significant increase in the correlation between the selection battery and the criterion.

To be able to convincingly demonstrate the feasibility of enhancing selection utility through the use of predictability indices would require the cross validation of the results obtained on a derivation sample on a holdout sample selected from the same population. The following two vital issues are at stake. The predictability index, developed on the derivation sample should still correlate significantly with the real, algebraic residuals obtained from fitting a new basic regression model on a representative holdout sample taken from the same population. Furthermore, the addition of the predictability index, developed on the derivation sample, to the holdout regression model should still significantly explain unique variance in the criterion measure that is not explained by the predictor(s) in the basic model. The first aspect is probably the Achilles heel of the proposed procedure. If the predictability index developed on the derivation sample would succeed in predicting the real prediction errors made by a newly fitted regression model on a second sample taken from the same population, then the second issue most likely will not present a problem. This study failed to investigate these two rather crucial aspects due to the limited size of the data set it had at its disposal.

TABLE 10
INCREMENTAL UTILITY AS A FUNCTION OF $R(Y, E[Y|X_1, X_2])$ AND SD_y

Na	T	Salary	Percent	$r(Y, X_1)$	$R(Y, E[Y X_1, X_2])$	C_1	C_2	λ	ϕ	Utility
100,00	5,00	100000,00	0,40	0,42	0,62	250,00	350,00	0,103	0,05	8443400,00
100,00	5,00	100000,00	0,40	0,42	0,72	250,00	350,00	0,103	0,05	12563400,00
100,00	5,00	100000,00	0,40	0,42	0,82	250,00	350,00	0,103	0,05	16683400,00
100,00	5,00	100000,00	0,40	0,42	0,92	250,00	350,00	0,103	0,05	20803400,00
100,00	5,00	100000,00	0,40	0,42	0,52	250,00	350,00	0,103	0,05	4323400,00
100,00	5,00	100000,00	0,40	0,42	0,42	250,00	350,00	0,103	0,05	203400,00
100,00	5,00	100000,00	0,40	0,42	0,42	250,00	350,00	0,103	0,05	-85000,00
100,00	5,00	100000,00	0,35	0,42	0,62	250,00	350,00	0,103	0,05	7377350,000
100,00	5,00	100000,00	0,35	0,42	0,72	250,00	350,00	0,103	0,05	10982350,00
100,00	5,00	100000,00	0,35	0,42	0,82	250,00	350,00	0,103	0,05	14587350,00
100,00	5,00	100000,00	0,35	0,42	0,92	250,00	350,00	0,103	0,05	18192350,00
100,00	5,00	100000,00	0,35	0,42	0,52	250,00	350,00	0,103	0,05	3772350,000
100,00	5,00	100000,00	0,35	0,42	0,42	250,00	350,00	0,103	0,05	167350,00
100,00	5,00	100000,00	0,35	0,42	0,42	250,00	350,00	0,103	0,05	-85000,0000
100,00	5,00	100000,00	0,45	0,42	0,62	250,00	350,00	0,103	0,05	9509450,000
100,00	5,00	100000,00	0,45	0,42	0,72	250,00	350,00	0,103	0,05	14144450,00
100,00	5,00	100000,00	0,45	0,42	0,82	250,00	350,00	0,103	0,05	18779450,00
100,00	5,00	100000,00	0,45	0,42	0,92	250,00	350,00	0,103	0,05	23414450,00
100,00	5,00	100000,00	0,45	0,42	0,52	250,00	350,00	0,103	0,05	4874450,00
100,00	5,00	100000,00	0,45	0,42	0,42	250,00	350,00	0,103	0,05	239450,000
100,00	5,00	100000,00	0,45	0,42	0,42	250,00	350,00	0,103	0,05	-85000,00

There is, moreover, a related question, which this study also failed to investigate. More in line with traditional cross validation of regression equations the question also arises to what extent the expanded regression model developed on the derivation sample would accurately predict the criterion when applied on the holdout sample data. In terms of the eventual regular use of predictability indices in selection research this clearly is an important issue.

The possibility of using bootstrapping to solve the problem of finding large enough initial samples to allow the division into derivation and holdout samples should be considered (Diaconis & Efron, 1983; Efron, 1982; Efron & Tibshirani, 1993). This procedure seems to present a feasible way of investigating the first two issues mentioned above. Whether it presents a solution to the more traditional cross validation problem seems somewhat more debatable.

Predictability indices most likely are highly situation specific. Each prediction model would most likely require the development of a unique predictability index. The fact that it was possible a predictability index for one prediction model does not necessarily mean it would practically be possible to do so for another. The question, therefore, also arises how common the occurrence of successful predictability index development actually is? Moreover it is not clear whether any criteria should be set for the type of donor predictor that would increase the likelihood of finding suitable donor items, and if so, what these criteria should be?

REFERENCES

- Abdi, H., Valentin, D. & Edelman, B. (1999). Neural networks. *Sage University Papers Series on Quantitative Applications in the Social Sciences*, No. 124. Thousand Oaks, CA: Sage.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Anderson, J.A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press
- Berenson, M.L, Levine, D.M & Goldstein, M. (1983). *Intermediate statistical methods and applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Binning, J.F. & Barrett, G.V. (1989). Validity of personnel decisions: a conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74, 478-494.
- Boudreau, J.W. (1991). Utility analysis for decisions in human resource management. In M.D. Dunnette & L.M. Hough (Eds.) *Handbook of industrial and organizational psychology* (2nd ed.) (pp. 621-745). Palo Alto, CA: Consulting Psychologists Press,
- Brogden, H.E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Education and Psychology*, 37, 65-76.
- Brogden, H.E. (1949a). When testing pays off. *Personnel Psychology*, 2, 171-185.
- Brogden, H.E. (1949b). A new coefficient: application to biserial correlation and to estimation of selective efficiency. *Psychometrika*, 14, 169-182.
- Campbell, J.P. (1991). Modelling the performance prediction problem in industrial and organisational psychology. In M.D. Dunnette & L.M. Hough (Eds), *Handbook of Industrial and Organizational Psychology* (pp. 39-74). Palo Alto, CA: Consulting Psychologists Press.
- Cascio, W.F. & Ramos, R.A. (1986). Development and application of a new method for assessing job performance in behavioral/economic terms. *Journal of Applied Psychology*, 71, 20-28.
- Cascio, W.F. (1991). *Costing human resources: the financial impact of behavior in organizations* (3rd ed.). Boston, MA: PWS-Kent.
- Cochran, W.G. (1951). Improvement by means of selection. In J. Neyman (Ed.), *Proceedings of the second Berkeley symposium on mathematical statistics and probability* (pp.449-470). Berkeley, CA: University of California Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cook, M. (1998). *Personnel selection: adding value through people* (3rd ed.). New York: Wiley
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.
- Cronbach, L.J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper & Row.
- Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, Ill: University of Illinois Press.
- Dawes, R.M. & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-106.
- Dawes, R.M. (1971). A case study of graduate admissions: application of three principles of human decision making. *American Psychologist*, 26, 180-188.
- Diaconis, P.M. & Efron, B.S. (1983). Computer-intensive methods in statistics. *Scientific American*, 116-130.
- Efron, B.S. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Efron, B.S. & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Gatewood, R.B. & Feild, H.S. (1994). *Human resource selection* (3rd ed.). Fort Worth, TX: Dryden.
- Ghiselli, E.E. (1956). Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology*, 40, 374-377.
- Ghiselli, E.E. (1960a). The prediction of predictability. *Educational and Psychological Measurement*, 20, 3-8.
- Ghiselli, E.E. (1960b). Differentiation of tests in terms of the accuracy with which they predict for a given individual. *Educational and Psychological Measurement*, 20, 675-84.
- Ghiselli, E.E. (1963). Moderating effects and differential reliability and validity. *Journal of Applied Psychology*, 47, 81-86
- Ghiselli, E.E., Campbell, J.P. & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco, CA: Freeman.
- Goldberg, L.R. (1970). Man versus model of man: a rationale plus evidence for a method of improving on clinical inferences. *Psychological Bulletin*, 73, 422-432.
- Grove, W.M. & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical- statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.
- Guion, R.M. (1991). Personnel assessment, selection and placement. In M. D. Dunnette & L. M. Hough (Eds), *Handbook of industrial and organizational psychology* (2nd ed.) (pp. 327-398). Palo Alto, CA: Consulting Psychologists Press.
- Guion, R.M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah, New Jersey: Lawrence Erlbaum.
- Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1995). *Multivariate data analysis with readings* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Howell, D. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, CA: Duxbury.
- Hull, C.L. (1928). *Aptitude testing*. London: Harrap.
- Kaplan, R.M. & Saccuzzo, D.P. (2001). *Psychological testing: principles, applications, and issues* (5th ed.). Belmont, CA: Wadsworth.
- Kerlinger, F.N., & Lee, H.B. (2000). *Foundations of behavioural research*. (4th ed.). Fort Worth, TX: Harcourt.
- Kleinmutz, B. (1990). Why we still use our heads instead of formulas: towards an integrative approach. *Psychological Bulletin*, 107, 296-310.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley
- March, J.G. & Simon, H.A. (1958). *Organizations*. New York: Wiley
- Meehl, P.E. (1954). *Clinical vs. statistical prediction: a theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meehl, P.E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology*, 4, 268-273.

- Milkovich, G.T. & Boudreau, J.W. (1994). *Personnel/human resources management: a diagnostic approach* (5th ed.). Homewood, Illinois: Richard Irwin Inc.
- Murphy, K.R. and Davidshofer, C.O. (1988). *Psychological testing: principles and applications*. Englewood Cliffs, NJ: Prentice Hall.
- Naylor, J.C. & Shine, L.C. (1965). A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 3, 33-42.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Psytech (2003). *Instruments*. World Wide Web (<http://www.psytech.co.za>). Psytech SA (Pty) Ltd.
- Saunders, D.R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement*, 16, 209-222.
- Schepers, J.M. (1996). The development of a statistical procedure to correct the effects of restriction of range on validity coefficients. *SA Journal of Industrial Psychology*, 22, 19-28.
- Schmidt, F.L. & Hunter, J.E. (1983). Individual differences in productivity: an empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology*, 68, 407-414.
- Smith, M. (1993). *Neural networks for statistical modelling*. New York: Van Nostrand Reinhold.
- Tabachnick, B.G. & Fidell, L.S. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper Collins.
- Taylor, H.C. & Russel, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Taylor, T. (1994). A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. *South African Journal of Psychology*, 24, 184-193.
- Theron, C.C. (1999). Psychometric implications of corrections for attenuation and restriction of range for selection validation research. Unpublished doctoral thesis, University of Stellenbosch.
- Theron, C.C. (2001). *The effect of a joint correction for the attenuating effect of criterion unreliability and case 1 restriction of range on the expected criterion performance and the conditional probability of success*. Paper presented at The 6th Conference of the European Association of Psychological Assessment, Aachen.
- Thorndike, R.L. (1982). *Applied psychometrics*. Boston, MA: Houghton- Mifflin.
- Tredoux, C.G. & Durrheim, K.L. (2002). *Numbers, hypotheses and conclusions*. Cape Town: Juta.
- Wiggins, J.S. (1973). *Personality and prediction: principles of personality assessment*. Reading, MA: Addison-Wesley.